DUPLICATE ALSO

The Met.Office

Forecasting Research

Forecasting Research Division
Scientific Paper No. 51

(*Submitted to Physics of Fluids*)

The Breakdown of Large-scale Flows in Rotating, Stratified Fluids

by

J C McWilliams, I Yavneh, M J P Cullen and P R Gent

May 1998

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

ORGS UKMO F

National Meteorological Library
FitzRoy Road, Exeter, Devon. EX1 3PB

The Breakdown of Large-scale Flows in Rotating, Stratified Fluids

James C. McWilliams*, Irad Yavneh†, Michael J.P. Cullen‡, and Peter R. Gent§

May 1998

Abstract

Flows under the influences of environmental rotation and stable density stratification often exhibit an approximate force balance and a consequently slow rate of evolution at large Reynolds number. Such flows are typically anisotropic in their velocity field. This regime is relevant to large-scale motions in the Earth's atmosphere and ocean, as well as many other planetary and astrophysical systems. The Balance Equations are usually an accurate approximate model for this regime. However, they have solvability limits associated with a change of type of their time-integration operator. In this paper we derive these limiting conditions for the conservative Balance Equations in isentropic coordinates, show that the least familiar of these conditions coincides with loss of convexity of the streamfunction for horizontal velocity in the inertial reference frame, and identify these conditions with the general conditions for symmetric loss of stability for circular and parallel flows as well as for the three-dimensional loss of stability for elliptical flows. We then conjecture that the identified limits of balance coincide generally with the boundary between the distinctive nonlinear dynamical behaviors (i.e., their turbulent cascade and dissipation

*Institute of Geophysics and Planetary Physics, UCLA, Los Angeles, CA, U.S.A
†Computer Science Department, Technion—Israel Institute of Technology, Haifa, Israel
‡Meteorological Office, Bracknell, England
§National Center for Atmospheric Research, Boulder, CO, U.S.A.

rates) associated with the large- and small-scale regimes in geophysical and astrophysical flows.

1 Introduction

Among the fundamental paradigms of theoretical fluid dynamics are the different advective behaviors of uniform-density, isotropic, homogeneous flows in two and three dimensions, 2D and 3D. In inviscid 2D flows, there are two quadratic integral invariants, kinetic energy and enstrophy (vorticity variance). In viscous flows with large Reynolds number, the energy is nearly conserved through its preferential transfer towards larger spatial scales (the inverse cascade), whereas the enstrophy and any passive tracer variance are dissipated through their preferential transfer towards smaller scales where viscous and diffusive effects are significant (the forward cascade). In 3D flows, in contrast, the kinetic energy is the only inviscid quadratic integral invariant, and the turbulent cascade to dissipation occurs primarily in a forward sense for energy, enstrophy, and tracer variance. Also, the forward cascade rate is much more efficient in 3D than in 2D. One measure of this is the finite predictability time in 3D as the initial error scale and Reynolds number, Re, both go to infinity, versus the infinite predictability time in 2D (Vallis, 1985; Metais and Lesieur, 1986). As another measure, the time for onset of dissipation, given smooth initial conditions, is finite and independent of Re in 3D, but it is a logarithmically increasing function of Re in 2D (Tatsumi and Yanase, 1981; also see Lesieur, 1990, pp. 185 and 281); this is related to the proof that bounded solutions exist for all time in 2D even with zero viscosity (Ebin and Marsden, 1970), whereas present computational evidence suggests singularities develop in a finite time in 3D (Kerr, 1993).

Fluid motions in the Earth's atmosphere and ocean, on scales comparable, respectively, to the depths of the troposphere and pycnocline or larger, are usually strongly anisotropic, with vertical velocity much smaller than horizontal. This is because the environmental influences of the planetary rotation and stable density stratification act to suppress vertical motions. Thus, in this sense, these motions are closer to the 2D paradigm than the 3D one. This is demonstrated in the asymptotic theory for 3D flows with strong rotation and stratification, quasigeostrophy (QG), where there are again two inviscid

quadratic integral invariants, total energy (kinetic plus available potential) and potential enstrophy (potential vorticity variance), which are associated, respectively, with inverse and forward cascades (Charney, 1971), as in 2D flows. On very small scales, however, flows are not strongly influenced by rotation and stratification, and the behavior is much closer to the 3D paradigm, with more nearly isotropic cascades towards dissipation at very small scales with the molecular diffusivities of air and water.

Much of the forcing of atmospheric and oceanic motions occurs on large scales, e.g., through planetary-scale differential heating or gyre-scale surface wind stress. In equilibrium kinetic energy and enstrophy budgets, the source terms due to large-scale forcing and conversions with other components of energy and potential enstrophy are typically comparable to the molecular dissipation by small-scale motions. Thus, there are flows of energy and enstrophy from the larger scales to smaller ones, which we can identify generically with the forward turbulent cascades mentioned above. This requires that there be energy and enstrophy flows from highly anisotropic, quasi-2D motions to nearly isotropic, 3D motions, and this necessarily involves a shift from the slower forward cascade of the former to the more rapid forward cascade of the latter. This transition we call the *breakdown* of the large-scale flows. In nature, of course, the cascades and dissipation occur quite inhomogeneously (e.g., especially in the planetary boundary layers), but this is not something we will consider here.

The Balance Equations (BE), as originally formulated by Lorenz (1960) and Charney (1962), are an approximate model for the anisotropic, rotating, stratified fluid motions on larger scales. This class of dynamics is sometimes loosely referred to as the slow manifold, since it typically has relatively little energy in the fast inertia-gravity wave modes permitted by the primitive fluid equations. Since BE dynamically exclude all fast waves, they are by construction a model only for evolution on the slow manifold and thereby give a precise meaning to what is otherwise a fuzzy concept.[1] The BE have one higher order of accuracy than QG asymptotically for the canonical regime of comparably strong rotation and stratification (i.e., equally small Rossby and Froude numbers, $Ro = V/fL$ and $Fr = V/NH$, where V is a character-

[1]There is an emerging consensus, but as yet no proof, that the primitive equations in general do not have an exact invariant manifold on which all evolution is slow, even when the Rossby number is small (Spall and McWilliams, 1992; Ford et al., 1998).

istic velocity, H and L are vertical and horizontal scales, and f and N are Coriolis and buoyancy frequencies) and a small ratio of characteristic vertical and horizontal scales. However, they also have second-order asymptotic validity in other anisotropic situations with planetary horizontal scale, with either strong rotation or stratification alone, and with flow configurations that are close to either frontal or vortical (Gent and McWilliams, 1983a; McWilliams, 1985; Gent et al., 1994; Shapiro and Montgomery, 1993). They also seem usually to be quite accurate even when the relevant order parameters (e.g., the Rossby number) are not particularly small (Norton et al., 1986; Allen et al., 1990a, 1990b; Barth et al., 1990; Whitaker, 1993; Yavneh and McWilliams, 1994). There have been many alternative formulations of balanced models, all of which have some form of hydrostatic and gradient-wind balance as the basis for approximating the vertical momentum equation and the vertical component of the curl and divergence of the horizontal momentum equations (Gent and McWilliams, 1983b; Gent and McWilliams, 1984; Allen, 1991; Holm, 1996; Vallis, 1996; McIntyre and Norton, 1998).[2] It is not yet clear whether any of these alternative balanced models is a universally superior one, since their solutions have been found to be quite similar in the few comparisons that have been made. Thus, there remains an unresolved ambiguity in the formulation of the BE at an asymptotically third order (e.g., Rossby number squared compared to the leading-order QG terms).

Gent and McWilliams (1984) argue in favor of the particular formulation that contains the most severely truncated approximation to the horizontal divergence equation consistent with retaining gradient-wind balance as expressed in isentropic coordinates (i.e., a transformation of the physical-space coordinates replacing height with potential density). The primary virtues of this formulation are its mathematical conciseness and its exact conservation of Ertel potential vorticity on fluid parcels in the absence of forcing or dissipation. The principal weakness of an isentropic formulation is the difficulty of obtaining reliable numerical solutions under extreme variations of

[2]There is also a large literature of other intermediate models (e.g., McWilliams and Gent, 1980, presents a now-dated review) that are not consistent with a second-order asymptotic approximation of the gradient-wind balance, and which therefore are less broadly accurate. The most famous of these, perhaps, is the Semigeostrophic Equations (Hoskins and Bretherton, 1972; Hoskins, 1975; Salmon, 1985), originally conceived for parallel frontal flows. Its balance relation has only leading order accuracy for non-parallel flows (Gent et al., 1994).

the stratification strength, where the coordinate transformation approaches a singularity; however, since this is also a condition of singularity of the BE more generally (see below), the isentropic BE are not in this way relatively disadvantaged compared to balanced models in physical-space coordinates.

Yavneh et al. (1997) analyze the solvability of the BE as formulated by Lorenz (1960) in physical-space coordinates. This shows that there are singularities in the time integrability of the BE that occur when any of several quantities lose positivity anywhere in the domain. We can therefore call these critical conditions the *limits of balance*. In this paper we reconsider this solvability analysis for the isentropic BE, both for the reasons given above and because the critical conditions have an even simpler mathematical form than in the physical-coordinate formulation. We then show that the critical condition for gradient-wind balance is equivalent to marginal convexity of the streamlines in an inertial (i.e., non-rotating) reference frame. We also show that the critical conditions for the limits of balance coincide exactly with the known critical conditions for the onset of symmetric instabilities in parallel and axisymmetric flows. Finally, we present the conjecture, along with some supporting evidence, that the limits of balance may generally coincide with the vigorous onset of nearly isotropic, and necessarily unbalanced, motions and the initiation of efficient 3D turbulent cascades to dissipation. This conjecture, therefore, identifies the limits of balance with the necessary transition from quasi-2D to quasi-3D dynamics in the atmospheric and oceanic cascades of energy and enstrophy.

2 Singularities in the Balance Equations

Let x, y and z denote Cartesian physical coordinates in the zonal, meridional, and vertical directions, respectively, and let t denote the time coordinate. Then, the isentropic coordinates are $X = x$, $Y = y$, and $Z = \theta(x, y, z, t)$, and the time coordinate is $T = t$. Here θ is the buoyancy field, defined as $-g\rho/\rho_o$, where ρ is the potential density, ρ_o is a reference value, and g is the gravitational acceleration constant. For simplicity we make the Boussinesq and incompressible approximations for the fluid.

The inviscid BE in isentropic coordinates (Gent and McWilliams, 1984) can be written as

$$D_T A + A\nabla^2\chi = 0\,, \tag{1}$$

$$D_T\,\phi_{ZZ} + \phi_{ZZ}\nabla^2\chi = 0\,, \tag{2}$$

$$\nabla^2\phi - \nabla\cdot(f\nabla\psi) - 2J(\psi_X,\psi_Y) = 0\,, \tag{3}$$

where subscripts X, Y or Z denote partial derivatives. Here, ψ is the horizontal streamfunction, $A = f + \zeta$ is the absolute vorticity (f denoting the Coriolis frequency, and $\zeta = \nabla^2\psi$ the "vertical" vorticity associated with the horizontal velocities), ϕ is the Montgomery potential (i.e., $p/\rho_o - z\theta$), χ is the velocity potential for the horizontally divergent component of the velocity, J is the horizontal Jacobian operator, and D_T is the material derivative given by

$$D_T \equiv \partial_T + J(\psi,\ ) + \nabla\chi\cdot\nabla\ .$$

All operators are defined with respect to the isentropic coordinates, and ∇ refers only to the horizontal gradient. Note that we have decomposed the horizontal velocity $\mathbf{v}_h$ into its rotational and divergent components, $(-\psi_Y,\ \psi_X)$ and $(\chi_X,\ \chi_Y)$.

We will restrict our attention to regions away from the equator (without loss of generality we assume $f > 0$), and balanced flow regimes that analytically continue from the QG limit. Thus, a Rossby number, $Ro = |\zeta|/f$, is a relevant control parameter, and our focus is on Ro values increasing from zero.

Eqns. (1), (2), and (3) are, respectively, the vorticity equation, the Montgomery-potential equation, and the divergence (gradient-wind balance) equation. Only in (3) is there a departure from the hydrostatic Primitive Equations in that the velocity-potential terms have been deleted. Eqns. (1) and (2) can be combined to show that Ertel potential vorticity is conserved following the flow,

$$D_T\,q \ = \ 0\,,$$

where $q \equiv AS$ and $S \equiv -\phi_{ZZ}^{-1}$.

A fundamental property of system (1-3) is that it is first order in time, as can be seen by taking the time derivative of (3) and then eliminating either ϕ_T or ψ_T and χ between (1) and (2). This differs from the temporally third order system of the more fundamental incompressible equations, where the degrees of freedom discarded by the BE correspond to (fast) inertia-gravity modes. Hence, in particular, only a single independent initial field can be chosen, and the remaining fields are tied to this field diagnostically. For example, given a sufficiently smooth ψ field we can readily compute the corresponding ϕ

and ζ fields by solving horizontal Poisson equations (given suitable boundary conditions, which are not discussed here.) It is then straightforward to derive the following linear system of partial differential equations for the χ, ϕ_T, and ψ_T fields from (1), (2), and the time-derivative of (3):

$$\begin{pmatrix} F & \partial_{ZZ} & 0 \\ 0 & \nabla^2 & -M \\ G & 0 & \nabla^2 \end{pmatrix} \begin{pmatrix} \chi \\ \phi_T \\ \psi_T \end{pmatrix} = \begin{pmatrix} -J(\psi, \phi_{ZZ}) \\ 0 \\ -J(\psi, A) \end{pmatrix}, \qquad (4)$$
$$\mathbf{L} \qquad\qquad \mathbf{u} \quad = \quad \mathbf{r}$$

where $\mathbf{L}$ has been written in matrix form for later convenience. Here, F, M and G are symmetric linear partial differential operators given by

$$\begin{aligned} F &= \nabla \cdot (\phi_{ZZ} \nabla) \\ M &= \nabla \cdot (f \nabla) + 2(\psi_{XX}\partial_{YY} + \psi_{YY}\partial_{XX} - 2\psi_{XY}\partial_{XY}) \\ G &= \nabla \cdot (A \nabla) . \end{aligned}$$

Below we shall analyze the operator $\mathbf{L}$, following the procedure of Volevič (1963) as described in Wloka et. al. (1995), where it is also shown to be equivalent to that of Douglis and Nirenberg (1955). It is simpler and illustrative to first analyze $\mathbf{L}$ in the f-plane QG limit by removing terms that depend on X and Y. We obtain

$$\begin{aligned} \mathbf{L}_{QG} &= \begin{pmatrix} \overline{\phi}_{ZZ}\nabla^2 & \partial_{ZZ} & 0 \\ 0 & \nabla^2 & -f_o\nabla^2 \\ f_o\nabla^2 & 0 & \nabla^2 \end{pmatrix} \\ &= \begin{pmatrix} \overline{\phi}_{ZZ} & \partial_{ZZ} & 0 \\ 0 & \nabla^2 & -f_o \\ f_o & 0 & 1 \end{pmatrix} \begin{pmatrix} \nabla^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \nabla^2 \end{pmatrix}, \end{aligned}$$

where f_o is a constant and $\overline{\phi}_{ZZ}$ depends only on Z. To examine the first factor of $\mathbf{L}_{QG}$, we follow the procedure of analyzing its characteristic polynomial. Taking the determinant of this operator we obtain

$$P_{QG}(\partial_X, \partial_Y, \partial_Z) = \overline{\phi}_{ZZ}\nabla^2 - f_o^2\partial_{ZZ} .$$

We only need to consider the principal part of P_{QG}, i.e., the highest-order terms, but here both terms are second-order. The operator is elliptic if, for

all real (k, ℓ, m),

$$\hat{P}_{QG}(k, \ell, m) \neq 0\,, \quad unless\ k = \ell = m = 0\,.$$

$\hat{P}$ is defined as the polynomial factor that results from applying P to $exp(kX + \ell Y + mZ)$. Here we have $\hat{P}_{QG}(k, \ell, m) = \overline{\phi}_{ZZ}(k^2+\ell^2) - f_o^2 m^2$. So the operator is evidently elliptic when $\overline{\phi}_{ZZ} < 0$ (stable stratification). If $\overline{\phi}_{ZZ}$ is positive in any region, the operator becomes hyperbolic, and characteristic surfaces exist, along which arbitrary nonsmoothness can occur; when $\overline{\phi}_{ZZ}$ is constant, these surfaces are cones. The boundary-value problem that is well-posed in the elliptic regime is then ill-posed.

The second term of $\mathbf{L}$ can be seen simply as a transformation of variables. Thus, one can first solve the system for the variables $(\nabla^2\chi, \phi_T, \nabla^2\psi_T)$, and later obtain χ and ψ_T. But $\mathbf{L}_{QG}$ itself is not elliptic, because its characteristic polynomial is $(k^2 + \ell^2)^2 \hat{P}_{QG}(k, \ell, m)$, which is equal to zero when $k = \ell = 0$. Such operators, which are elliptic with respect to some surface (and for which the characteristic polynomial does not change sign but may vanish), are generally called degenerate-elliptic operators (see, e.g., Morioka, 1991; Horiuchi, 1994; Levendorskii, 1990). (In Yavneh et. al., 1997, and elsewhere, this is referred to as semi-elliptic, but this is not the most common usage of this term).

We now examine $\mathbf{L}$ similarly. The determinant of $\mathbf{L}$ can formally be written as

$$P(\partial_X, \partial_Y, \partial_Z) = F\nabla^4 - \partial_{ZZ} MG\,. \tag{5}$$

Now, retaining only the highest-order terms we obtain

$$\hat{P}(k, \ell, m) = (k^2 + \ell^2)\ [\ \phi_{ZZ}(k^2 + \ell^2)^2 - m^2 A\hat{M}(k, l)\ ]\,, \tag{6}$$

where

$$\hat{M} = f(k^2 + \ell^2) + 2(\psi_{XX}\ell^2 + \psi_{YY}k^2 - 2\psi_{XY}k\ell)\,.$$

Again, evidently, $\hat{P} = 0$ for $k = \ell = 0$. This results from the fact that the horizontal curl and divergence operators were applied when deriving the BE. Hence, we again seek conditions under which $\mathbf{L}$ is degenerate-elliptic. This requires a sign-definite $\hat{P}$ for any positive $k^2 + \ell^2$. In the limit of stable stratification (negative definite ϕ_{ZZ}) and strong rotation ($f >> |\nabla^2\psi|$), this clearly holds since $\hat{P} \to (k^2+\ell^2)^2[-\phi_{ZZ}(k^2+\ell^2) + f^2m^2] > 0$. But degenerate-ellipticity is lost in any of the following three cases:

1. Change of sign of $S \equiv -\phi_{ZZ}^{-1}$;

2. Change of sign of A;

3. Change of sign of $\hat{M}$.

It is straightforward to show (Yavneh et. al., 1997) that $\hat{M}$ is negative definite (hence, M, and the Monge Ampere operator from which it is derived are elliptic), if and only if

$$A^2 - \sigma^2 > 0\,, \tag{7}$$

where σ is the magnitude of the strain rate associated with the horizontal, rotational part of the flow:

$$\sigma = [\ (\psi_{XX} - \psi_{YY})^2 + 4(\psi_{XY})^2\ ]^{1/2}\,. \tag{8}$$

Evidently, change of sign of $\hat{M}$ precedes change of sign of A as Ro increases (or coincides with it in the special case where σ and A vanish simultaneously).

We are not aware of any proof that (4), given appropriate boundary conditions, constitutes a well-posed problem (as in the QG limit above), but there is much numerical evidence that supports this presumption. However, when one of the three conditions above is violated, the system becomes hyperbolic, and then the problem certainly is no longer well-posed. Hence, we consider these to be necessary solvability conditions.

Note that these solvability conditions are instantaneous ones. They must be satisfied by the initial conditions and at all later times for the continuing integrability of the BE. However, there is no assurance *a priori* that violations cannot develop in time without non-conservative forcing or damping.[3]

The solvability analysis of the BE formulated in physical-space coordinates (Yavneh et. al., 1997) identified three critical conditions analogous to those above. The only differences are the following: (i) a physical-coordinate BE singularity occurs slightly before the first sign change due to an asymptotically higher order term that enters into the positivity condition; (ii) the strain rate and vorticity quantities which appear in the critical conditions have slightly different meanings because horizontal gradients are different in the two coordinate systems; and (iii) the critical conditions cannot be cleanly

[3]In contrast, solvability violations cannot develop in time in the conservative Semi-geostrophic Equations, given favorable boundary conditions (Cullen and Purser, 1989).

expressed in terms of q in the physical-coordinate BE. Nevertheless, since we believe the solution behaviors for these two different BE formulations are similar under most circumstances, it is plausible that the differences in their singularity conditions will be neither large nor dynamically significant.

2.1 Special cases

We now consider certain restrictions of our problem to special cases, in particular some cases for which there are known stability criteria for the Boussinesq equations.

Stratified Barotropic flow. If we apply the restriction that all derivatives with respect to Z vanish for $\mathbf{v}_h$ and $\nabla\phi$, then (4) becomes triangular (by exchanging the orders of both the variables ϕ_T and ψ_T and their associated equations), and it is clearly degenerate-elliptic so long as ϕ_{ZZ} is sign-definite. Thus, the second and third conditions are only relevant for non-barotropic motions.

Parallel flow If we assume the flow is parallel to the X coordinate, say, by setting all derivatives with respect to Y to zero, we obtain by (5), $M = \nabla\cdot(f\nabla)$, which is uniformly elliptic with respect to the (X,Y) plane. Thus, the third condition for degenerate-ellipticity does not apply when streamlines (including any perturbations) are restricted to be parallel straight lines. The first two conditions remain, i.e.,

$$A = f + \psi_{YY} > 0, \qquad S > 0.$$

Axisymmetric flow If the streamlines (and any perturbations) are restricted to be axisymmetric, then M is given in isentropic polar coordinates by

$$M = \frac{1}{r}\partial_r(fr\partial_r) + \frac{2}{r}\partial_r(\psi_r\partial_r) = \frac{2}{r}\partial_r[\,(\frac{1}{2}fr + \psi_r)\partial_r\,]\,. \tag{9}$$

Thus, in the axisymmetric case the third condition of degenerate-ellipticity is sign-definiteness of the absolute circulation, $C \equiv \frac{1}{2}fr + \psi_r$, so the three critical conditions can here be expressed as

$$fC > 0, \quad A = f + \frac{1}{r}\partial_r(r\psi_r) > 0, \quad S > 0\,. \tag{10}$$

3 Convexity in the Inertial Frame

It is interesting to consider the second and third criteria for degenerate-ellipticity from the point of view of the inertial reference frame, instead of the rotating frame. The preceding equations apply when the axis of rotation is vertical and the angular velocity of the rotation is $f/2 = \Omega$, where Ω is the local (in latitude) projection of the angular rotation of the Earth. Thus, the transformation between rotating $(X,\ Y)$, and non-rotating, $(\xi,\ \eta)$, horizontal coordinates is given by

$$\begin{aligned} X &= \xi \cos[ft/2] + \eta \sin[ft/2] \\ Y &= -\xi \sin[ft/2] + \eta \cos[ft/2]. \end{aligned}$$

We denote the streamfunction in the non-rotating frame by Ψ. It is related to the rotating-frame streamfunction by

$$\Psi = \psi + \frac{1}{4} f r^2,$$

where $r^2 = X^2 + Y^2 = \xi^2 + \eta^2$. We can further show that the nonlinear operator in (3) is invariant under the rotational transformation, i.e.,

$$J_{XY}(a_X, a_Y) = J_{\xi\eta}(\mathcal{A}_\xi, \mathcal{A}_\eta)$$

for any $a(X,Y) \equiv \mathcal{A}(\xi,\eta)$.

We can rewrite the third critical condition (7) as

$$\frac{1}{4} f^2 + \frac{1}{2} f(\psi_{XX} + \psi_{YY}) + J_{XY}(\psi_X, \psi_Y) > 0,$$

or, in inertial-frame variables,

$$J_{\xi\eta}(\Psi_\xi, \Psi_\eta) = \Psi_{\xi\xi}\Psi_{\eta\eta} - \Psi_{\xi\eta}^2 > 0\,. \tag{11}$$

Eqn. (11) is often referred to as the convexity condition, since it implies that the curvature of the $\Psi(\xi,\eta)$ function is everywhere of one sign. Thus, the critical condition related to the Monge-Ampere operator in gradient-wind balance is equivalent to convexity of the streamfunction surface in the inertial reference frame.[4]

[4]Note that the convexity of the streamlines required for the Semigeostrophic Equations is with respect to a frame rotating with angular velocity $f/\sqrt{2}$, which is not the Earth's inertial frame (Cullen and Purser, 1989); however, the axisymmetric version of the Semigeostrophic Equations does give the correct frame (Shutts et al., 1988).

Convexity of a function is a very strong constraint on its behavior. It severely restricts small scale oscillations. The derivative of a convex function has to vary monotonically in any given direction. If a sequence of convex functions converges to a limit, then their derivatives also converge to the derivative of the limit function. The only oscillations permitted are in the second derivatives. These properties are very important in showing that solutions of certain partial differential equations are well-posed, and giving conditions under which the discrete numerical approximations converge to them. In physical applications, convexity occurs naturally in stability theory. If the energy is a convex (or concave) function within a neighborhood of a given state, then stability of that state is guaranteed (Thomson, 1910). It has also been more recently exploited by Shutts and Cullen (1987) in their analysis of parcel stability in rotating stratified flows. In steady flows, convexity of the streamfunction prevents exponential growth of line elements, and only permits algebraic growth. Convexity has also proved useful in studies of two-dimensional turbulence: coherent structures and oscillation of scalar gradients are associated with convex regions of the streamfunction, and rapidly straining regions where scalar gradients grow are associated with areas where the streamfunction has saddle points (e.g., Weiss, 1991).[5]

4 Symmetric Instability Boundaries

Symmetric instability is defined as the instability of a symmetric mean flow to perturbations that have the same symmetry. Here we review what is known about the necessary and sufficient conditions for symmetric instability, in relation to the special cases of the limits of balance analyzed above.

For barotropic mean flows in a stratified fluid, barotropic instabilities can occur, but they do not violate the limits of balance. Indeed, they can occur even within the QG approximation. Thus, this type of instability remains on the slow manifold and can be accurately calculated with the BE.

For parallel mean flows, the condition for symmetric instability is a change of sign of potential vorticity, $q = AS$ (Hoskins, 1974). This is identical to the critical conditions for limits of balance in this case. When this condition

[5]Since the vorticity equation governing the dynamics of a two-dimensional flow is unaltered by uniform rotation, this characterization refers to the component of the streamfunction that is independent of the reference frame.

is violated the fastest growing linear instabilities have small spatial scale in both X and Z. (Stone, 1966)[6]

For axisymmetric instability, the conditions for symmetric instability are the change of sign of either the absolute circulation C or the Ertel potential vorticity, $q = AS$ (Ooyama, 1966). Again these conditions coincide with the limits of balance in this case, and the fastest growing linear instabilities are small-scale in both r and Z. (Smyth and McWilliams, 1998).[7]

Therefore, the known boundaries for onset of symmetric instabilities, which result in highly unbalanced growing modes, occur exactly at the limits of balance for these symmetric regimes in the BE formulated in isentropic coordinates.

5 Elliptical Instability Boundaries

Elliptical instability can arise in steady, vertically uniform, horizontal flows with a spatially uniform vorticity and with streamlines that are ellipses in the horizontal plane. This is a well studied problem for 3D instabilities in a non-rotating, unstratified fluid since its discovery by Pierrehumbert (1986), Craik and Criminale (1986), and Bayly (1986); it has also been analyzed for the rotating, stably stratified regime by Miyazaki (1993).

We now consider the relation between elliptical instability and the limits of balance as derived above. The instability of this flow to convective motions for $S < 0$ is unsurprising. For elliptical streamlines, the positivity of $\hat{M}$ is a more severe constraint than the positivity of A. McWilliams and Yavneh (1998) shows that, although there are some weakly unstable, exponentially growing modes within the regime of validity of the BE, the fluctuation growth rates become dramatically larger as soon as $\hat{M}$ becomes negative; furthermore, this instability mode has very small spatial scales, suggestive of the onset of a nearly isotropic 3D cascade.

[6]The Semigeostrophic Equations give the same answers as the BE about the coincidence of solvability limits and symmetric instability for parallel flow (Shutts and Cullen, 1987).

[7]An axisymmetric analog of the Semigeostrophic Equations gives the same answers as the BE about the coincidence of solvability limits and symmetric instability for axisymmetric flow (Shutts et al., 1988), but no single intermediate model of the Semigeostrophic class correctly characterizes both parallel and axisymmetric flows. An extension of the model by Shutts et al. (1988) for almost axisymmetric flows is made by Cullen and Purser (1989).

6 Isotropic, Three-dimensional Cascades to Dissipation

We conjecture that the behavior which occurs in symmetric flows is likely to be generally true for asymmetric, unsteady balanced flows: when the general conditions for the degenerate-ellipticity of (4) are violated, a breakdown of the slow manifold will occur, with vigorous growth of unbalanced, nearly isotropic 3D motions.[8] The violation may arise spontaneously through a balanced evolution (which is precluded in Semigeostrophic dynamics) or through the effects of non-conservative forcing or damping. The fluctuation growth will occur initially by linearized dynamics, but it will continue until nonlinear dynamics converts it into a rapid 3D cascade to dissipation. This breakdown will no doubt sometimes be accompanied by inertia-gravity wave radiation.[9] We further conjecture that this behavior will be spatially localized in the vicinity of the violation of the critical conditions; if so, then after the breakdown and cascade to dissipation of the unbalanced fluctuations, the larger-scale fields will have been locally altered by the associated mixing, and the flow may then be able to return to a slower balanced evolution. Also, we conjecture that the fastest growing modes associated with this phenomenon are likely to be of small spatial scales, as these are, of course, the scales that are dominant in the principal part of the operator. This indeed occurs in the special symmetric and elliptic instability cases noted above. As we have not been able to prove that $\mathbf{L}$ is sign-definite when the degenerate-ellipticity conditions are satisfied (only its principal part), there may exist growth modes even before violation occurs, but we expect these to be large-scale. This is unlikely to occur in practice, even if such modes do potentially exist, because regions of near-violation will typically be localized in space.

These conjectures are well supported for the first critical condition by

[8] Because of the ambiguity in the definition of balanced models discussed in the introduction, there is an analogous ambiguity in the identification of the limits of balance.

[9] In a shallow-water example, the unbalanced wave component was found to be small compared to the unbalanced cascading component in a solution that closely approached the only condition for the limits of balance, i.e., vanishing layer depth (Yavneh and McWilliams, 1994). However, this situation is not quite analogous to the 3D BE, because its limiting condition is the same as the necessary condition for existence of the shallow-water equations themselves and thus cannot be violated while there still exists a valid dynamics more fundamental than the BE.

many demonstrations of the outbreak of convection when the stratification becomes unstable (i.e., $S < 0$). The significance of the second critical condition is less well tested, but asymmetric, unbalanced, unstable modes have been found for some flows that locally violate the condition $A > 0$ (Lesieur et al., 1991). However, the third critical condition (7), related to the inertial-frame convexity of the flow, has as yet been explored less extensively. But it has been shown to be an important regime boundary for 3D fluctuations on a stratified, barotropic, balanced mean flow with spatially uniform strain rate and vorticity, in that the unbalanced fluctuations grow much more rapidly when it is violated (McWilliams and Yavneh, 1998).

In summary, the limits of balance are often, and perhaps usually, indicative of the dynamical regime boundary between the distinctive behaviors of large-scale, anisotropic and small-scale, nearly isotropic motions in geophysical and astrophysical rotating, stratified flows.

7 Acknowledgements

This research was supported in part by the National Science Foundation (through the National Center for Atmospheric Research), the Office of Naval Research (grant N00014-98-1-0165), the U.S.–Israel Bilateral Science Foundation (grant 94-00250), and the U.K. Meteorological Office.

References

[1] Allen, J.S., Barth, J.A., and Newberger, P.A., 1990a: On intermediate models for barotropic continental shelf and slope flow fields. Part I: Formulation and comparison of exact solutions. *J. Phys. Oceanogr.* **20**, 1017-1042.

[2] Allen, J.S., Barth, J.A., and Newberger, P.A., 1990b: On intermediate models for barotropic continental shelf and slope flow fields. Part III: Comparison of numerical model solutions in periodic channels. *J. Phys. Oceanogr.* **20**, 1949-1973.

[3] Allen, J.S., 1991: Balance equations based on momentum equations with global invariants of potential enstrophy and energy. *J. Phys. Oceanogr.* **21**, 265-276.

[4] Barth, J.A., J.S. Allen, and P.A. Newberger, 1990: On intermediate models for barotropic continental shelf and slope flow fields. Part II: Comparison on numerical model solutions in doubly periodic domains. *J. Phys. Oceanogr.* **20**, 1044-1076.

[5] Bayly, B.J., 1986: Three-dimensional instability of elliptical flow. *Phys. Rev. Lett.* **57**, 2160-2163.

[6] Charney, J. G., 1962: Integration of the primitive and balance equations. *Proc. Int. Symp. Numerical Weather Prediction*, Tokyo, Meteor. Soc. Japan, 131-152.

[7] Charney, J.G., 1971: Geostrophic turbulence. *J. Atmos. Sci.* **28**, 1087-1095.

[8] Craik, A.D.D., and W.O. Criminale, 1986: Evolution of wavelike disturbances in shear flow: a class of exact solutions of the Navier-Stokes equations. *Proc. R. Soc. London Ser. A* **406**, 13-***.

[9] Cullen, M.J.P., and R.J. Purser, 1989: Properties of the Lagrangian semigeostrophic equations. *J. Atmos. Sci.* **46**, 2684-2697.

[10] Douglis, A., and L. Nirenberg, 1955: Interior estimates for elliptic systems of partial differential equations. *Comm. Pure Appl. Math.* **8**, 503-538.

[11] Ebin, D., and J. Marsden, 1970: Groups of diffeomorphisms and the motion of an incompressible fluid. *Annals Maths.* **92**, 102-163.

[12] Ford, R., M. McIntyre, and W.A. Norton, 1998: Balance and the slow quasi-manifold: some explicit results. *J. Atmos. Sci.*, submitted.

[13] Gent, P.R., and J.C. McWilliams, 1983a: Regimes of validity for balanced models. *Dyn. Atmos. Oceans* **7**, 167–183.

[14] Gent, P.R., and J.C. McWilliams, 1983b: Consistent balanced models in bounded and periodic domains. *Dyn. Atmos. Oceans* **7**, 67–93.

[15] Gent, P.R., and J.C. McWilliams, 1984: Balanced models in isentropic coordinates and the shallow water equations. *Tellus* **36A**, 166–171.

[16] Gent, P.R., J.C. McWilliams, and C. Snyder, 1994: A note on a scaling analysis of curved fronts: the formal validity of the balance equations and semigeostrophy. *J. Atmos. Sci.* **51**, 160-163.

[17] Holm, D., 1996: Hamiltonian Balance Equations. *Physica D*, **98**, 379-414.

[18] Horiuchi, T., 1994: Notes on the mean value property for certain degenerate elliptic operators. *J. Math. Soc. Japan* **46**, 149-163.

[19] Hoskins, B.J., and F.P. Bretherton, 1972: Atmospheric frontogenesis models: mathematical formulation and solution. *J. Atmos. Sci.* **29**, 11-37.

[20] Hoskins, B.J., 1974: The role of potential vorticity in symmetric stability and instability. *Q. J. Royal Met. Soc.* **100**, 480-482.

[21] Hoskins, B.J., 1975: The geostrophic momentum approximation and the semigeostrophic equations. *J. Atmos. Sci.* **32**, 233-242.

[22] Kerr, R.M., 1993: Evidence for a singularity of the three-dimensional, incompressible Euler Equation. *Phys. Fluids A* **5**, 1725-1746.

[23] Lesieur, M., 1990: *Turbulence in Fluids* (2nd edition), Kluwer, Dordrecht, 412 p.

[24] Lesieur, M., S. Yanase, and O. Metais, 1991: Stabilizing and destabilizing effects of solid body rotation on quasi-two-dimensional shear layers. *Phys. Fluids A* **3**, 403-407.

[25] Levendorskii, S.Z., 1990: On types of degenerate elliptic operators. *Math. of the USSR – Sbornik* **66** (2) 523-540.

[26] Lorenz, E.N., 1960: Energy and numerical weather prediction. *Tellus* **12**, 364-373.

[27] McIntyre, M.E., and W. Norton, 1998: Potential vorticity inversion on a hemisphere. To be submitted.

[28] McWilliams, J.C., and P.R. Gent, 1980: Intermediate models of planetary circulations in the atmosphere and ocean. *J. Atmos. Sci.* **37**, 1657–1678.

[29] McWilliams, J.C., 1985: A note on a uniformly valid model spanning the regimes of geostrophic and isotropic, stratified turbulence: balanced turbulence. *J. Atmos. Sci.* **42**, 1773–1774.

[30] McWilliams, J.C. and I. Yavneh, 1998: Fluctuation growth and instability associated with a singularity of the balance equations. *Physics of Fluids*, in press.

[31] Metais, O., and M. Lesieur, 1986: Statistical predictability of decaying turbulence. *J. Atmos. Sci.* **43**, 857-870.

[32] Miyazaki, T., 1993: Elliptical instability in a stably stratified rotating fluid. *Phys. Fluids A* **5** 2702-2709.

[33] Morioka, T., 1991: Hypoellipticity for semi-elliptic operators which degenerate on hypersurface. *Osaka J. Math.* **28**, 563-578.

[34] Norton, N.J., J.C. McWilliams, and P.R. Gent, 1986: A numerical model of the Balance Equations in a periodic domain and an example of balanced turbulence. *J. Comp. Phys.* **67**, 439–471.

[35] Ooyama, K., 1966: On the stability of the baroclinic circular vortex: a sufficient condition for instability. *J. Atmos. Sci.* **23**, 43-53.

[36] Pierrehumbert, R.T., 1986: Universal short-wave instability of two-dimensional eddies in an inviscid fluid. *Phys. Rev. Lett.* **57**, 2157-2159.

[37] Salmon, R., 1985: New equations for nearly geostrophic flow. *J. Fluid Mech.* **153**, 461-477.

[38] Shapiro, L.J., and M.T. Montgomery, 1993: A three-dimensional balance theory for rapidly rotating vortices. *J. Atmos. Sci.* **50**, 3322-3335.

[39] Shutts, G.J., and M.J.P. Cullen, 1987: Parcel stability and its relation to semigeostrophic theory. *J. Atmos. Sci.* **44**, 1318-1330.

[40] Shutts, G.J., M. Booth, and J. Norbury, 1988: A geometric model of balanced, axisymmetric flows with embedded penetrative convection. *J. Atmos. Sci.* **45**, 2609-2621.

[41] Smyth, W.D. and J.C. McWilliams, 1998: Stability of an axisymmetric vortex in a stably stratified rotating environment. *Theor. and Comp. Fluid Dyn.*, in press.

[42] Spall, M.A., and J.C. McWilliams, 1992: Rotational and gravitational influences on the degree of balance in the shallow-water equations. *Geophys. Astrophys. Fluid Dyn.* **64**, 1-30.

[43] Stone, P.H., 1966: On non-geostrophic baroclinic instability. *J. Atmos. Sci.* **23**, 390-400.

[44] Tatsumi, T., and S. Yanase, 1981: The modified cumulant expansion for two-dimensional isotropic turbulence. *J. Fluid. Mech.* **100**, 475-496.

[45] Thomson, W. (Lord Kelvin), 1910: Maximum and minimum energy in vortex motion. *Math. and Phys. Papers* **4**, 172-183. (Cambridge University Press)

[46] Vallis, G.K., 1985: Remarks on the predictability properties of two- and three-dimensional flow. *Q. J. Roy. Met. Soc.* **111**, 1039-1047.

[47] Vallis, G.K., 1996: Potential vorticity inversion and balanced equations of motion for rotating and stratified flows. *Q. J. Roy. Met. Soc.* **122**, 291-322.

[48] Volevič, L.R., 1963: A problem in linear programming coming from differential equations. *Usp. Mat. Nauk* **18** (3) 155-162.

[49] Weiss, J., 1991: The dynamics of enstrophy transfer in two-dimensional turbulence. *Physica D* **48**, 273-294.

[50] Whitaker, J.S., 1993: A comparison of primitive and balance equation simulations of baroclinic waves. *J. Atmos. Sci.* **50**, 1519-1530.

[51] Wloka, J.T., B. Rowley, and B. Lawruk, 1995: *Boundary value problems for elliptic systems.* Cambridge University Press, New York.

[52] Yavneh, I., and J.C. McWilliams, 1994: Breakdown of the slow manifold in the Shallow-Water Equations. *Geophys. Astrophys. Fluid Dyn.* **75**, 131-161.

[53] Yavneh, I., A.F. Shchepetkin, J.C. McWilliams, and L.P. Graves, 1997: Multigrid solution of rotating, stably stratified flows: The balance equations and their turbulent dynamics. *J. Comp. Physics* **136**, 245-262.